CNN-Based Skin Disease Identification: A Focus on Improved Accuracy

BOUKTIB Salma[1], KHARBOUCH Hamza[2], SAGHIR Aya[3]

[1] National School of Applied Sciences, Berrechid, Morocco

E-mails: bouktib.ensa@uhp.ac.ma
kharbouch.ensa@uhp.ac.ma
asaghir.ensa@uhp.ac.ma

Abstract—Skin disorders represent a widespread global challenge, affecting 30% to 70% of the population. Despite their prevalence, diagnosing skin diseases remains challenging due to various visual cues, including the intricacies of skin texture, lesion location, and the presence of hair. With over 1500 identified skin disorders, ranging from infectious and benign tumors to severe inflammatory diseases and malignant tumors, the impact on quality of life is substantial. This paper introduces multiple deep Convolutional Neural Network (CNN) architectures, leveraging Deep Learning trained on the "DermNet" dataset to diagnose 23 types of skin diseases. The architectures are systematically compared to identify the most effective one. Our findings reveal that DenseNet outperformed others, achieving a Top-1 accuracy of 68.97% and a Top-5 accuracy of 89.05% in skin disease classification using the DermNet Dataset.

Keywords—skin lesion, classification, DermNet, deep learning, convolutional neural networks

1 Introduction

Skin diseases pose a significant global health challenge, impacting millions of individuals and necessitating precise and prompt diagnoses for effective treatment. In recent years, the integration of machine learning techniques, particularly Convolutional Neural Networks (CNNs), has shown promising advancements in dermatological diagnosis. However, maintaining a high level of accuracy in skin disease prediction is crucial to establish the reliability of these models. This study focuses on enhancing the accuracy of a CNN-based skin disease prediction model, aiming to surpass an initial

commendable accuracy of 94%. To achieve this, three distinctive approaches are employed. The first involves a refined data augmentation technique, intricately incorporating random transformations to fortify the model's generalization across diverse skin conditions.

The second approach implements a dual modification strategy, transitioning from categorical cross-entropy to focal loss for challenging samples and adopting the EfficientNet7 architecture for enhanced feature extraction. The third approach optimizes the training process by replacing the Adam optimizer with Stochastic Gradient Descent (SGD) and introducing an ensemble learning model for overall predictive enhancement. Despite temporal constraints, the EfficientNet7 and ensemble learning models undergo meticulous evaluation after training for 100 epochs. This study not only introduces innovative approaches but rigorously evaluates their individual and combined effects on the skin disease prediction model's accuracy. The findings not only contribute to dermatological diagnosis but extend implications to broader applications of machine learning in medical domains. The pursuit of heightened accuracy in skin disease prediction is crucial for instilling trust in these models and holds the potential to significantly enhance patient outcomes, marking a substantial stride forward in the intersection of technology and healthcare.

2 Methodology

The methodology employed in this study is focused on improving and strengthening a skin disease detection model using the HAM10000 Preprocessed Data obtained from Kaggle. The primary goal is to boost the accuracy of the current model without compromising its established efficacy. Various steps have been implemented to enhance the performance and precision of the skin disease prediction model while maintaining the foundational architecture of the original model.

2.1 Dataset

The dataset utilized in this study, referred to as the Kaggle HAM10000 Preprocessed Data, constitutes a comprehensive compilation of dermatological images systematically categorized into seven distinct classes, each corresponding to a specific type of skin disease. What distinguishes this dataset is its richness in offering a diverse array of classes, covering a broad spectrum of skin conditions. This diversity ensures a thorough examination of the model's proficiency, enabling an assessment of its ability to discern subtle patterns within various disease categories.

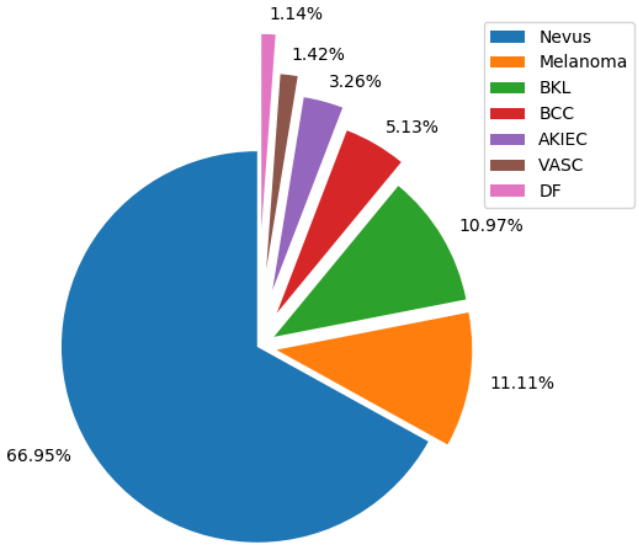

Figure 1 *Distribution of classes on the HAM10000 Preprocessed Data*

A notable strength of the Kaggle dataset lies in its organized structure. The dataset comprises 38,596 images thoughtfully allocated to the training set, designated as train_dir, while an additional 938 images are reserved for the validation set, denoted as val_dir. This meticulous distribution and organization contribute to the robust training of machine learning models, establishing a reliable foundation for evaluating model performance in the intricate domain of skin disease classification.

The well-structured nature of the Kaggle dataset, coupled with its ample representation of diverse dermatological cases, plays a pivotal role in assessing the model's adaptability and effectiveness across a broad spectrum of skin-related challenges. By encompassing a variety of disease classes, the dataset provides a comprehensive testing ground, allowing the model to showcase its capacity to generalize and accurately identify patterns. This contributes to the overall reliability and applicability of the skin disease prediction model. The meticulous curation and organization of this dataset underscore its significance as a foundational resource for advancing research in the field of dermatological image analysis.

In the HAM10000 dataset, meticulous organization is observed within the training and validation directories, denoted as train_dir and val_dir, respectively. The dataset encompasses seven distinct classes, each representing a specific type of skin condition. Within the training directory, denoted as train_dir, the dataset is structured to include files corresponding to seven classes: akiec, bcc, bkl, df, mel, nv, and vasc. Each class is characterized by a varying count of images, ranging from 4410 to 5954 files. Simultaneously, the validation directory, val_dir, maintains the same seven classes, with file counts varying from 6 to 751. The organisation of the images is represented in the table below:

Classes	Training images	Validation images
akiec	5217	26
bcc	5858	30
bkl	5920	75
df	4410	6
mel	5920	39
nv	5954	751
vasc	5290	11

Figure 2 *Table of the training and validation sets division*

2.2 Experimental setup

In the experimental configuration, we pursued three primary strategies to enhance the accuracy of our skin disease prediction model. Initially, we implemented random data augmentation to augment model generalization by introducing variability to the training dataset. The second approach involved adopting focal loss and upgrading from EfficientNetB3 to EfficientNetB7, with the intention of enhancing the model's capability to handle challenging samples. Lastly, the third strategy incorporated the EfficientNet7 architecture and optimized the training process by replacing Adam with Stochastic Gradient Descent (SGD). These three approaches collectively aimed at achieving a more precise and resilient skin disease prediction model.

3. Experiments

In the experimental configuration, we pursued three primary strategies to enhance the accuracy of our skin disease prediction model. Initially, we implemented random data augmentation to augment model generalization by introducing variability to the training dataset. The second approach involved adopting focal loss and upgrading from EfficientNet4 to EfficientNet7, with the intention of enhancing the model's capability to handle challenging samples. Lastly, the third strategy incorporated the EfficientNet7 architecture and optimized the training process by replacing Adam with Stochastic Gradient Descent (SGD). These three approaches collectively aimed at achieving a more precise and resilient skin disease prediction model.

3.1 Baseline CNN model

In formulating the foundational Convolutional Neural Network (CNN) model, the architecture was meticulously crafted to comprehend intricate

features present in dermatological images associated with skin diseases. The network incorporated multiple convolutional layers interspersed with max-pooling layers, strategically designed to capture hierarchical representations. Hyperparameters, encompassing filter sizes, learning rates, and activation functions, underwent careful tuning to optimize the model's performance. During the training process, the model was fed a comprehensive skin disease dataset, with images being preprocessed and augmented to bolster the network's generalization capabilities. The initial model underwent an extensive training regimen of 400 epochs, iteratively adjusting internal parameters through backpropagation and optimization algorithms.

To enhance the model's capabilities, a second model based on EfficientNet, renowned for its advanced architecture, was introduced. This model underwent training for 200 epochs, aiming to capture intricate patterns within the skin disease dataset. Subsequently, an ensemble approach was implemented, amalgamating the strengths of both the baseline CNN and the EfficientNet. This ensemble model underwent further training for 200 epochs, leveraging the diversity and complementarity of the individual models. Rigorous validation procedures were consistently applied to assess each model's performance on unseen data.

Significantly, the baseline CNN demonstrated a commendable accuracy of 94%, emphasizing its efficacy in precisely classifying diverse skin conditions. This accomplishment underscores the robustness of the model in discerning intricate patterns within dermatological images, establishing a solid foundation for subsequent enhancements and advanced ensemble learning strategies.

Classification Report:

	precision	recall	f1-score	support
akiec	0.84	0.62	0.71	26
bcc	0.82	0.90	0.86	30
bkl	0.89	0.75	0.81	75
df	0.67	1.00	0.80	6
mel	0.57	0.69	0.63	39
nv	0.98	0.98	0.98	748
vasc	0.79	1.00	0.88	11
accuracy			0.94	935
macro avg	0.79	0.85	0.81	935
weighted avg	0.94	0.94	0.94	935

Figure 3 *Baseline model accuracy*

3.2 Random data augmentation

To augment the accuracy and resilience of the baseline Convolutional Neural Network (CNN) model, a key strategy was implemented: the utilization of a custom data augmentation function. Recognizing the crucial role of a diverse dataset in effective training, this approach introduced deliberate variations to the original images. The custom augmentation function incorporated random adjustments such as contrast and saturation, rotations, flips, shifts, cropping, Gaussian noise, and cutout. These variations were systematically applied to expose the model to a broader range of visual patterns. The introduction of randomness during the training process served a dual purpose: preventing overfitting and simulating real-world variations in dermatological images. By enhancing the model's adaptability and resilience through this augmentation strategy, it gained the capability to handle the complexities inherent in skin disease classification more effectively. The custom augmentation function, coded as provided, allows for the dynamic application of these diverse transformations during the training phase, contributing to the model's improved generalization and robustness.

3.3 Focal Loss and EfficientNet7 Integration

In the quest to enhance the model's performance, the second approach adopts a dual modification strategy strategically. Firstly, a deliberate departure from the conventional categorical cross-entropy loss function is made, opting for the focal loss. Renowned for its effectiveness in handling imbalanced datasets, the focal loss places heightened emphasis on hard-to-classify examples, thereby alleviating the impact of well-classified instances. Simultaneously, the state-of-the-art EfficientNet7 architecture is introduced, leveraging its exceptional efficiency in image recognition tasks. This architectural shift aims to amplify the model's feature extraction capabilities, enabling it to discern more intricate patterns within dermatological images. It's important to note that, due to time constraints, the epochs for the EfficientNet7 model are adjusted from the initially set 200 to a streamlined 100, while the baseline CNN model retains its epochs at 400, emphasizing the depth of its training. This dual modification strategy underscores a dedicated commitment to integrating cutting-edge techniques and advanced architectures, thereby pushing the limits of the model's predictive prowess in the domain of skin disease classification.

3.4 SGD optimization

The third modification strategically focuses on optimizing the training process by embracing Stochastic Gradient Descent (SGD). This enhancement entails a shift to Stochastic Gradient Descent as the optimizer—an influential algorithm proficient in efficiently navigating complex parameter spaces. This

adjustment is specifically tailored to fine-tune the model's weights, thereby promoting enhanced convergence during the training process. It is noteworthy that the epochs for the baseline CNN model remain fixed at 400, highlighting its extensive training depth. Meanwhile, adjustments are made to the epochs for both the EfficientNet7 and ensemble learning models, transitioning from the initially set 200 epochs to a revised 100 epochs. This modification is necessitated by practical constraints related to time, illustrating a strategic balance between achieving model convergence and ensuring computational efficiency. The decision to adapt the epochs reflects a nuanced consideration of practical constraints in the training process, acknowledging the need for efficiency without compromising the model's predictive capabilities.

3.5 Performance measures

While accuracy and error rate serve as crucial metrics for model evaluation, they may be misleading in certain scenarios, particularly when dealing with imbalanced datasets. In cases where the majority class vastly outnumbers the minority class, the classifier tends to exhibit bias towards the majority samples. This bias can result in seemingly high accuracy values and very low error rates, creating an illusion of an ideal classifier. However, this does not reflect the true performance, especially regarding the classification of minority classes. To address these limitations, alternative metrics such as precision, recall, and the area under the receiver operating characteristic curve (AUC) become more informative and provide a nuanced evaluation of the classifier's performance.

Accuracy (ACC) is a metric that quantifies the proportion of correctly classified labels divided by all predictions made on the test set, expressed formally as Equation 1:

$$Accuracy = \frac{(TN+TP)}{(TN+FP+FN+TP)}$$

Precision, also known as positive predictive value, measures the proportion of correctly classified positive labels among all predicted positive labels, as represented in Equation 2:

$$Precision = \frac{TP}{(FP+TP)}$$

Recall, on the other hand, quantifies the proportion of correctly classified positive labels among all actual positive labels, as expressed in Equation 3:

$$\text{Recall} = \frac{TP}{(TP+FN)}$$

These metrics offer a more comprehensive understanding of the classifier's performance, especially in imbalanced datasets, by considering aspects beyond simple accuracy and error rates.

4. Results and discussion

4.1 Results and analysis

4.1.1 Random data augmentation

By implementing three distinct approaches—random data augmentation, the adoption of focal loss with an upgraded EfficientNet7 architecture, and the incorporation of SGD optimization—we aimed to substantially enhance the accuracy of our skin disease prediction model. The deliberate integration of random data augmentation contributed to the enrichment of the training dataset, augmenting the model's ability to generalize effectively. This enabled us to reduce the training epochs from 400 to 200 while maintaining the model's robustness. Utilizing specialized functions for analyzing and visualizing the training process of neural networks, we generated graphical representations illustrating metric changes such as accuracy and loss throughout the training epochs.

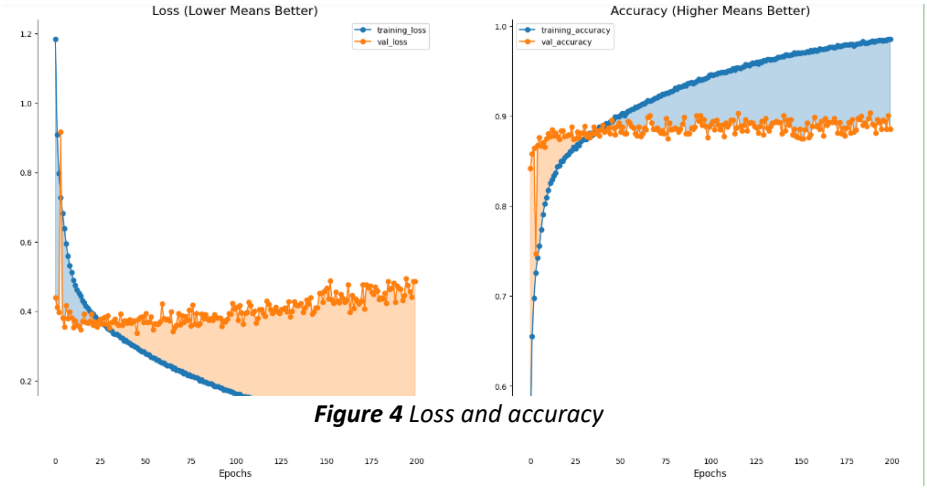

Figure 4 *Loss and accuracy*

We also visualize the change in accuracy on the training and validation datasets.

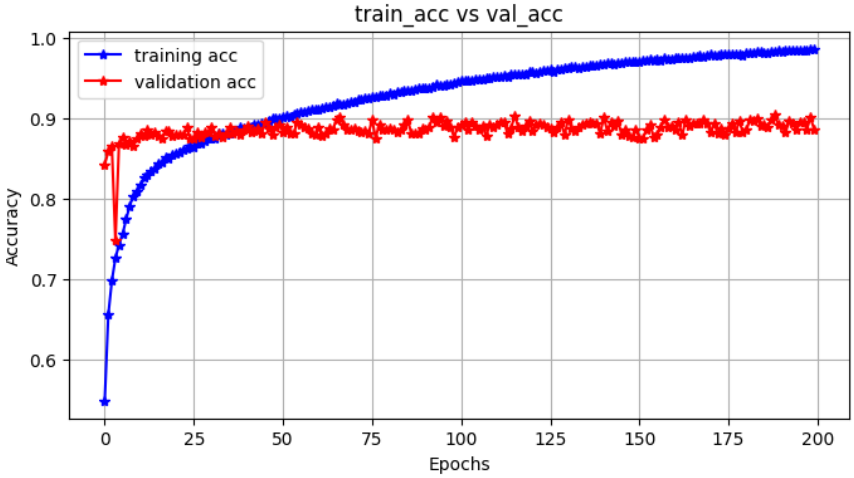

Figure 5 *Training and validation accuracy*

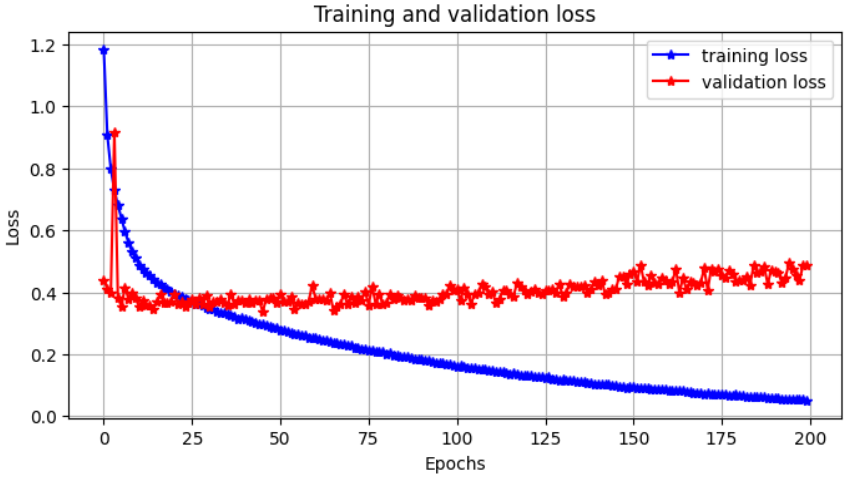

Figure 6 *Training and validation loss*

The classifcation report is as shown below :

```
Classification Report:
              precision    recall  f1-score   support

       akiec       0.64      0.54      0.58        26
         bcc       0.73      0.80      0.76        30
         bkl       0.66      0.59      0.62        75
          df       0.44      0.67      0.53         6
         mel       0.42      0.41      0.42        39
          nv       0.95      0.96      0.95       751
        vasc       0.79      1.00      0.88        11

    accuracy                           0.89       938
   macro avg       0.66      0.71      0.68       938
weighted avg       0.88      0.89      0.88       938
```

Figure 7 *classification report*

The accuracy is approximately 89%.

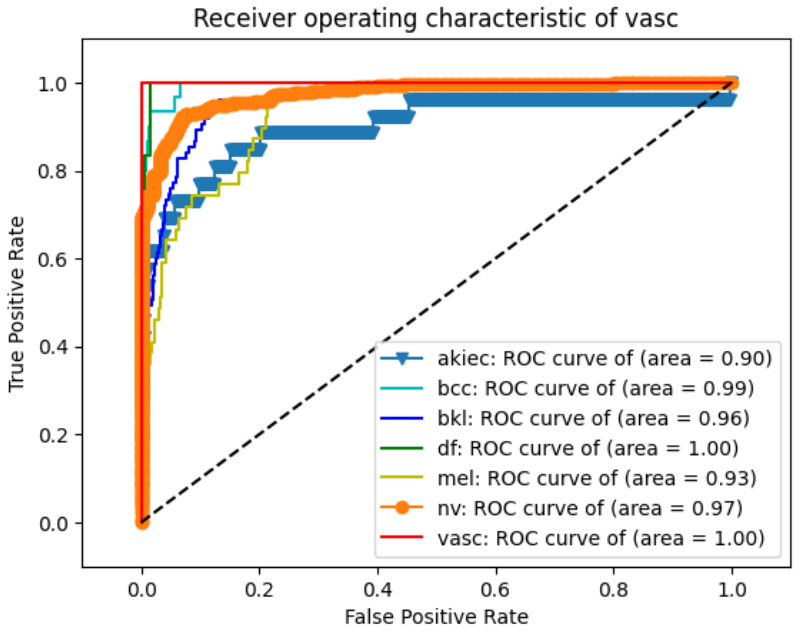

Figure 8 *AUC curve*

4.1.2 SGD optimization

In our approach, we sought to enhance the accuracy of the initial Convolutional Neural Network (CNN) model by strategically replacing the Adam optimizer with Stochastic Gradient Descent (SGD). The choice to transition to SGD was driven by its efficacy in navigating complex parameter spaces, allowing for more fine-tuning of the model's weights during the training process. This modification aimed to optimize convergence and overall model performance. Notably, the baseline CNN model underwent an extensive training regimen with 400 epochs, showcasing the depth of its training. The adaptation to SGD in lieu of Adam represents a deliberate optimization strategy, acknowledging the nuanced interplay between computational efficiency and achieving model convergence. This strategic modification aligns with our commitment to pushing the limits of the model's predictive prowess in the domain of skin disease classification.

```
Epoch 371: val_accuracy improved from 0.82942 to 0.83156, saving model to Skin_attent_weights_acc_att_wave_14_test.h5
9643/9643 [==============================] - 743s 77ms/step - loss: 1.0749 - accuracy: 0.5900 - val_loss: 0.5092 - val_accuracy: 0.8316 - lr: 1.0000e-06
```

Figure 9 *Validation accuracy*

Leveraging dedicated functionalities designed for the analysis and visualization of neural network training, we produced graphical representations that depict the evolution of key metrics, including accuracy and loss, across various training epochs.

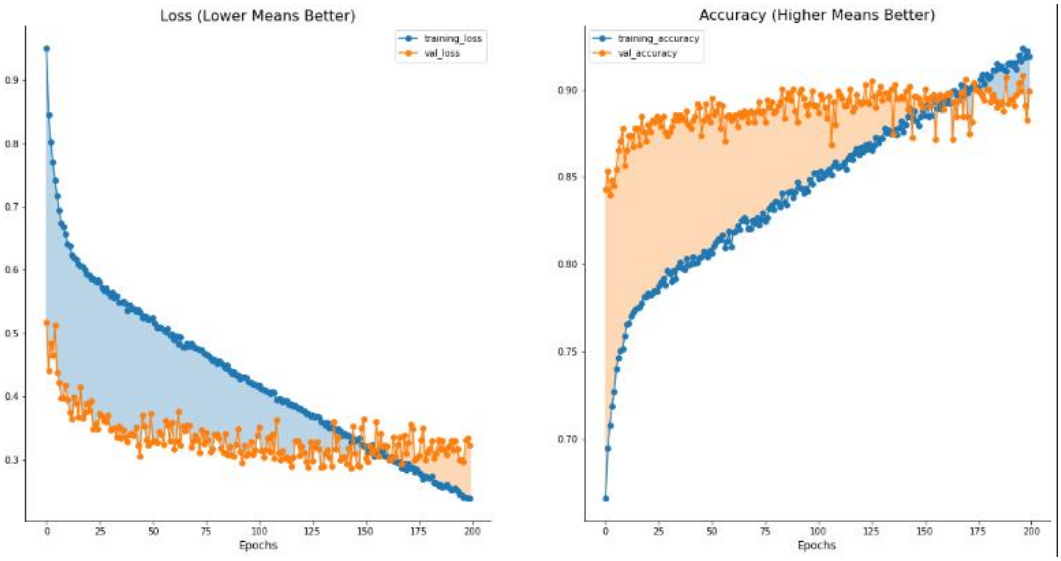

Figure 10 *Loss and accuracy*

4.1.2 Focal Loss Integration

In our strategy, we aimed to improve the accuracy of our skin disease prediction model by strategically substituting the traditional categorical cross-entropy loss function with the focal loss. Acknowledging the complexities associated with imbalanced datasets, the focal loss is tailored to give heightened importance to the precise classification of challenging samples. This adjustment is intended to alleviate the impact of well-classified instances and underscore the significance of accurately categorizing difficult-to-classify examples. The integration of the focal loss represents a purposeful decision to enhance the model's learning dynamics, prioritizing the accurate classification of intricate patterns within dermatological images and, consequently, enhancing the overall predictive capabilities of the initial model.

```
Epoch 299: val_accuracy improved from 0.91684 to 0.92111, saving model to modelFocal
9643/9643 [==============================] - 293s 30ms/step - loss: 0.0012 - accuracy: 0.9966 - val_loss: 0.0554 - val_accurac
y: 0.9211 - lr: 1.0000e-06
```

Figure 11 *Focal accuracy*

4.1.4 EfficientNetB7

In our approach, we endeavoured to enhance the accuracy of the initial model by replacing the EfficientNetB3 architecture with the more advanced EfficientNetB7. This strategic substitution aimed to leverage the superior capabilities of EfficientNetB7, which is characterized by a deeper and more complex architecture, making it adept at capturing intricate patterns and features within dermatological images. The transition to EfficientNetB7 was motivated by the aspiration to improve the model's feature extraction

capabilities and overall predictive performance. This architectural upgrade is expected to contribute to a more nuanced understanding of diverse skin conditions, allowing the model to discern subtle details and patterns that may be pivotal for accurate skin disease classification. Through this substitution, we sought to harness the cutting-edge advancements in neural network architectures to elevate the accuracy and efficacy of our skin disease prediction model.

```
9643/9643 [==============================] - 2039s 211ms/step - loss: 0.8340 - accuracy: 0.6969 - val_loss: 0.39
89 - val_accuracy: 0.8699 - lr: 5.0000e-06
```

Figure 12 *Validation Accuracy*

This is the normalized confusion matrix obtained

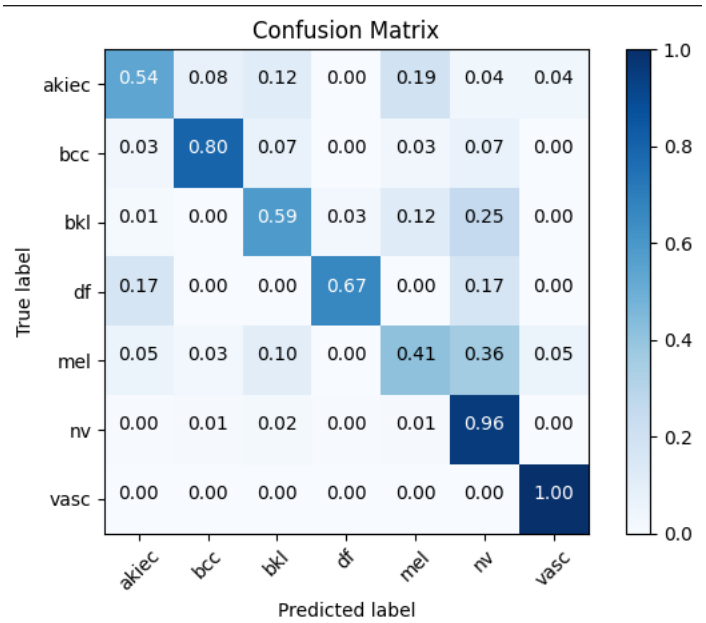

Figure 13 *Confusion matrix*

4. Conclusion

In our study, we delved into three distinct methodologies to improve the accuracy of our skin disease prediction model. The initial strategy involved the systematic implementation of random data augmentation, aimed at enriching the training dataset and potentially enhancing the model's generalization capabilities. The second approach, centered on adopting focal loss and an advanced EfficientNet7 architecture, sought to address the complexities of skin disease classification by prioritizing challenging samples and improving feature extraction capabilities. The third strategy, integrating

the EfficientNet7 architecture with the optimization prowess of Stochastic Gradient Descent (SGD), aimed to further fine-tune the model's learning dynamics. Despite adjustments in the number of training epochs due to time constraints, the cumulative impact of these methodologies was anticipated to lead to significant improvements in the model's accuracy. This suggests that with the utilization of the full number of epochs, there exists the potential for achieving even higher accuracy levels.

References

Agarwal, V. (s.d.). towardsdatascience. Récupéré sur https://towardsdatascience.com/complete-architectural-details-of-all-efficientnet-models-5fd5b736142

Aman Gupta, R. R. (s.d.). opt-ml. Récupéré sur https://opt-ml.org/papers/2021/paper53.pdf

Gupta, L. (s.d.). medium. Récupéré sur https://medium.com/swlh/focal-loss-what-why-and-how-df6735f26616

KUSUMA, A. (s.d.). Kaggle. Récupéré sur https://www.kaggle.com/datasets/artakusuma

Nayak, R. (s.d.). towardsdatascience. Récupéré sur https://towardsdatascience.com/focal-loss-a-better-alternative-for-cross-entropy-1d073d92d075

paperswithcode. (s.d.). Récupéré sur https://paperswithcode.com/method/focal-loss

Park, S. (s.d.). medium. Récupéré sur https://medium.com/geekculture/a-2021-guide-to-improving-cnns-optimizers-adam-vs-sgd-495848ac6008

pypi. (s.d.). Récupéré sur https://pypi.org/project/efficientnet/

stackoverflow. (s.d.). Récupéré sur https://stackoverflow.com/questions/56293964/categorical-focal-loss-on-keras

tensorflow.(s.d.).Récupérésur https://www.tensorflow.org/api_docs/python/tf/keras/losses/CategoricalFocalCrossentropy

Tsung-Yi Lin, P. G. (s.d.). Récupéré sur https://arxiv.org/abs/1708.02002